

Power-Efficient System Design for Cellular-Based Machine-to-Machine Communications

Harpreet S. Dhillon, Howard C. Huang, Harish Viswanathan and Reinaldo A. Valenzuela

Abstract—The growing popularity of Machine-to-Machine (M2M) communications in cellular networks is driving the need to optimize networks based on the characteristics of M2M, which are significantly different from the requirements that current networks are designed to meet. First, M2M requires large number of short sessions as opposed to small number of long lived sessions required by the human generated traffic. Second, M2M constitutes a number of battery operated devices that are static in locations such as basements and tunnels, and need to transmit at elevated powers compared to the traditional devices. Third, replacing or recharging batteries of such devices may not be feasible. All these differences highlight the importance of a systematic framework to study the power and energy optimal system design in the regime of interest for M2M, which is the main focus of this paper. For a variety of coordinated and uncoordinated transmission strategies, we derive results for the optimal transmit power, energy per bit, and the maximum load supported by the base station, leading to the following design guidelines: (i) frequency division multiple access (FDMA), including equal bandwidth allocation, is sum-power optimal in the asymptotically low spectral efficiency regime, (ii) while FDMA is the best practical strategy overall, uncoordinated code division multiple access (CDMA) is almost as good when the base station is lightly loaded, (iii) the value of optimization within FDMA is in general not significant in the regime of interest for M2M.

I. INTRODUCTION

The widespread coverage of cellular networks makes them an attractive option for handling the growing number of sensing and monitoring devices. Therefore, M2M communications, involving wide area communication of sensor data to an Internet based application, is emerging as an important service over mobile cellular networks [2]–[5]. It spans multiple vertical industries such as transportation, healthcare, utilities, retail, industrial monitoring, banking, and home automation and includes a variety of applications within each vertical. Projections for growth of M2M communication devices range from 24 billion [6] to 50 billion [7] in the next decade with over 2 billion M2M devices expected to directly attach to the cellular network by this time. Given the potential for a significant new revenue stream from M2M data services, the industry is focusing on ensuring that cellular networks can efficiently serve the needs of M2M communications.

A. Motivation and Related Work

M2M devices in some verticals are deployed in locations that are not frequented by people. For example, vending ma-

chines and water, gas or other meters, are typically deployed in basements of buildings, water monitoring systems are deployed underground, and some traffic monitoring systems may be deployed in the tunnels. Such machines will typically have devices that communicate with their controlling applications or servers over wireless networks. For such locations, radio signals have to be substantially stronger compared to what is required for traditional service. Hence, device transmit power required for communication becomes a critical issue. In addition, if the devices are battery operated total transmission energy from the device is also an important consideration.

For M2M, the goal of minimizing mobile transmit power is aided by the nature of M2M traffic. In contrast to traditional consumer traffic, M2M typically involves a large number of short payload transactions, as shown in [8], [9], where traffic models for various vertical applications are characterized. For example, a fleet management application can involve transmission of location every 20 seconds by each vehicle to the central application server with each transaction involving less than 500 bytes [10]. Similarly, reporting of health data such as blood pressure or heart rate by medical devices involves payloads less than 200 bytes [11].

Mobile cellular networks, including the fourth generation Long Term Evolution (LTE) [12] are neither designed with link budget requirements of M2M devices, nor optimized for M2M traffic pattern. The system design is optimized to maximize spectral efficiency and minimize latency. Thus mobiles transmit in short duration bursts at high power levels that maximize the total sector throughput. Mobile transmit power levels are primarily dictated by maximum transmit power limit of mobiles and out-of-cell interference considerations. Battery life is not a primary concern for communications since the dominant power consumption on human devices are driven by displays and complex application processing. Furthermore, users will recharge their devices as required.

Clearly, M2M communications impose new requirements on cellular networks that demand rethinking some of the design principles [13], [14]. First, for short transactions it may be advantageous to transmit payload in the random access request itself instead of establishing dedicated bearers [15]. This depends on the size of the payload, the overheads involved and the level of base station loading. Second, the optimization criterion for resource allocation and transmit power levels is average transmit power or energy consumed to transmit a given payload. This is because for M2M devices battery life is an important concern, and communications consumes a significant fraction of the battery energy, especially if the devices have to transmit at elevated powers due to their

H. S. Dhillon is with WNCG, the University of Texas at Austin, USA. Email: dhillon@utexas.edu. H. C. Huang, H. Viswanathan and R. A. Valenzuela are with Bell Labs, Alcatel-Lucent, NJ. A part of this paper was presented at IEEE Globecom Workshops 2012 [1]. Manuscript last revised: December 11, 2013.

adverse locations. Motivated by these differences, several modifications in the current communication protocols to reduce signaling overhead [15]–[17] and power consumption [18] have been proposed in the literature. Furthermore, the idea of cooperative design where several devices are clustered together with a possibility of a controller acting as a common link between a cellular base station and the devices is investigated, e.g., in [19], [20]. The problem of uplink scheduling of M2M devices in LTE networks is studied in [21], where it is shown that it is better not to form different classes of the devices in order to increase the maximum load that can be served at the base station. Despite these research efforts, there is little understanding of the fundamental tradeoffs in the parameter space of interest in M2M communications, especially from a power and energy optimal design perspective, which is the main focus of this paper.

B. Contributions

Evaluate different multiple access strategies and identify the minimum power strategy. We study both uncoordinated and coordinated multiple access strategies in this paper. In uncoordinated strategies, the payload is carried in the very first message together with the control information such as the device identity and thus there are no dedicated resources allocated. We consider both FDMA and CDMA random access strategies. The transmit power, number of frequency channels in FDMA and spread factor in CDMA are adjusted based on the average load on the system, which is known at the devices by listening to downlink broadcast signaling from the base station. Interested readers can refer to [22] for a comparison of the two schemes from a throughput perspective. We then consider coordinated strategies in which the resources are explicitly scheduled to active devices. The base station determines the transmission time duration, bandwidth and power and indicates this to the devices. Here we consider successive interference cancellation (SIC), FDMA and time division multiple access (TDMA) strategies and focus on determining the optimal bandwidth/time and transmit power settings for minimizing the total power and/or energy consumed to transmit a fixed size payload. Comparison of the uncoordinated and coordinated strategies shows, not surprisingly, that the coordinated strategies outperform the uncoordinated strategies for heavy loading whereas the performance is comparable for light loads. Hence for low loads, an uncoordinated strategy may be preferred when taking into account downlink overhead required to inform the devices of the allocated resources.

Bound the gap between optimal and equal-resource allocation in coordinated orthogonal transmission. Minimum energy and/or power scheduling over time or frequency is known to be a convex optimization problem and various efficient algorithms have been proposed in the literature [23]–[25]. We show that, in the cellular setting, it is possible to approximate the optimal resource allocation through simpler one-shot solutions, which leads to closed form expressions for the optimization parameters in certain special cases. Furthermore, we prove that in the parameter space of interest, somewhat surprisingly, simple equal-resource allocations perform within

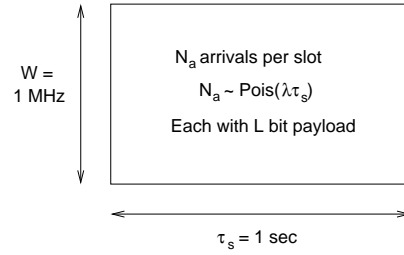


Fig. 1. Illustration of the time-frequency resource “slice” over which multiple users are scheduled.

a very small factor of the optimal allocations both in terms of transmit power and energy per bit for both TDMA and FDMA. The bounding factor is derived in closed form for all the cases and is around 1.25 for both FDMA sum-power and sum-energy and also for TDMA sum-energy. It is around 2 for TDMA sum-power. This suggests that a simple equal resource allocation algorithm is sufficient to achieve near-optimal transmission in terms of both the power and energy minimization.

Asymptotically optimal strategies for coordinated transmission. Using the closed form bounds derived to bound the gap between the optimal and equal resource allocation, we further show that the FDMA equal resource allocation is both sum-power and sum-energy optimal over the space of FDMA strategies in the limit of asymptotically low spectral efficiency. Since the closed form bound is the same for TDMA energy optimal solution, the result extends in this case as well. Using these bounds and the closed form result derived for the sum-power required in SIC, we also prove that FDMA, including equal resource allocation strategy, is sum-power optimal over general resource allocation strategies (not necessarily orthogonal) in the limit of asymptotically low spectral efficiency.

The rest of the paper is organized as follows. We introduce notation and describe the system model in Section II. We discuss the uncoordinated and coordinated access strategies and the associated system designs in Sections III and IV, respectively. We present the numerical results comparing different strategies in Section V. We present our conclusions in Section VI.

II. SYSTEM MODEL

A. System Setup

In this paper, we consider a single cell consisting of an access point lying at the origin and devices uniformly distributed around it in an annular region with inner and outer radii r_i and r_o , respectively. The non-zero inner radius is assumed to avoid singularity in the path loss model, which is discussed later in this section. We focus on a single cell system with no out-of-cell interference, which is one of the simulation scenarios in 3GPP model [13]. It should be noted that the out-of-cell interference effectively changes the operating signal-to-interference-plus-noise ratio (SINR), which can be incorporated in the current analysis to study the multi-cell case. In this study, we focus only on the uplink. To characterize the uplink

load seen by the base station, we model the arrival process of packets as a Poisson point process with mean λ arrivals per second. For concreteness, we assume a time slotted system with the slot duration denoted by τ_s . The analysis will be performed on a typical time-frequency resource slice with slot duration τ_s and bandwidth W as shown in Fig. 1. We denote the number of packet transmission requests in each block by $N_a \sim \text{Pois}(\lambda\tau_s)$. Each packet is assumed to have a payload of L bits.

B. Multiple Access

Uplink multiple access is primarily enabled by a broadcast or beacon signal that is transmitted by the base station at the start of each time slot. The beacon signal is assumed to carry load information, which is characterized by the average load λ seen at the base station. We consider two broad categories of multiple access strategies: i) uncoordinated: the devices transmit data using slotted random access and there is no need to establish dedicated radio bearers, and ii) coordinated: the devices transmit data in a separate scheduled transmission. An example of the system using uncoordinated strategy is the one-stage setup illustrated in Fig. 2, where both the control information and the data are transmitted as a part of the random access signal. As discussed in the sequel, the transmit power and other transmission parameters are independently adjusted by each mobile device based on the load information, λ , that is received in the beacon signal.

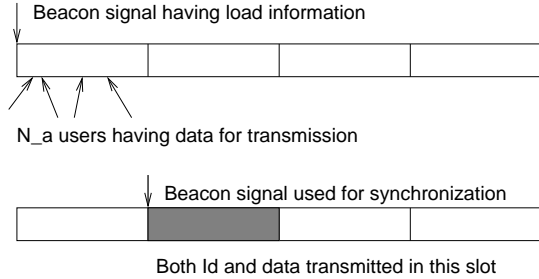


Fig. 2. One-stage design where both device Id and data are transmitted together in random access stage.

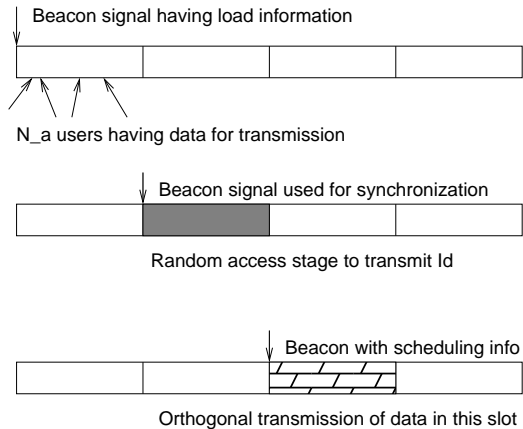


Fig. 3. Two-stage design where device Id is transmitted in the random access stage to establish connection followed by the scheduled data access stage.

An example of the system using a coordinated strategy is the two-stage setup shown in Fig. 3. In this setup, a dedicated uplink connection is first established by transmitting only the control information in the random access stage. The random access design is the same as the one discussed for the one-stage system. The base station then transmits scheduling information as a part of the beacon signal, which is used by the devices to transmit orthogonally from each other.

In all the strategies considered in this paper, we assume that all the packet transmissions occur in the current slot and none of them are left for scheduling in a future slot, thereby introducing a notion of a packet deadline explicitly. Furthermore, we only consider successful packet transmission in a single attempt since the packet failure rate will be small by design. The effective deadline for a single slot design is τ_s and for two-stage design is $2\tau_s$, counting time from the beginning of the first slot after the packet arrives.

C. Modeling Uplink Channel from Device to Base Station

The received power at the base station from a device located at distance r , assuming transmit power P_t , path loss exponent γ , small-scale fading gain h , large scale shadowing gain \mathcal{X} and direction based antenna gain G is:

$$P_r = P_t \mathcal{X} h G r^{-\gamma}. \quad (1)$$

Instead of treating these link budget parameters individually, we model their composite effect by defining the reference signal-to-noise ratio (SNR) μ as the average received SNR from a device transmitting at maximum power P_{\max} over bandwidth W located at cell edge, i.e., at distance r_0 from the base station. Therefore, the received SNR μ_r at the base station from a device located at distance r can be expressed in terms of μ as:

$$\mu_r = \frac{P_t}{P_{\max}} \mu \mathcal{X} h \left(\frac{r}{r_0} \right)^{-\gamma}. \quad (2)$$

Please note that reference SNR is a function of signal bandwidth because of the scaling of noise power with bandwidth. We will comment more on this when we study coordinated access using TDMA and FDMA in the sequel. Now defining the channel gain $g = \mathcal{X} h \left(\frac{r}{r_0} \right)^{-\gamma}$, we get the following simple expression for the received SNR μ_r :

$$\mu_r = \frac{P_t}{P_{\max}} \mu g. \quad (3)$$

In case the information symbols are transmitted over a bandwidth $W_{\mathcal{N}} < W$, the reference SNR $\mu_{\mathcal{N}}$ in this case can be written in terms of μ as:

$$\mu_{\mathcal{N}} = \mu \frac{W}{W_{\mathcal{N}}}, \quad (4)$$

which is greater than μ because of the decrease in the effective noise power. For this modified system, the received SNR can be evaluated from (3) by replacing μ with $\mu_{\mathcal{N}}$. This will be helpful in analyzing the multiple access strategies that involve partitioning of frequency resources, e.g., FDMA. Throughout this paper, we assume that the channel gain g is known at

the device both in the cases of uncoordinated and coordinated strategies.

For all numerical results, we consider $P_{\max} = 1$ W, $r_0 = 1000$ m, $\gamma = 3$, $\mu = -3$ dB, $W = 1$ MHz, $\tau_s = 1$ sec and ignore fading and shadowing. A careful reader would observe that the value of μ is very low for a reference distance of 1000 m. This is chosen to account for the high penetration losses suffered by the radio signals when the devices are deployed at adverse locations such as basements and tunnels. Since we have chosen $P_{\max} = 1$ W, we will drop it from (3) to reduce it to $\mu_r = P_t \mu g$ with the understanding that the P_t is the normalized by $P_{\max} = 1$ W.

D. Preliminaries

In the slotted system introduced above, an information symbol of payload L bits will be transmitted over bandwidth $W_N \leq W$ for time $\tau \leq \tau_s$ depending upon the resources allocated to that packet. To fix the key ideas, we first confine our discussion to the perfectly orthogonal resource allocation using TDMA and FDMA, where one of these inequalities will be a strict inequality due to the partitioning of time-frequency resources. Assuming capacity achieving codes, τ and W_N are related to the received SNR μ_r by Shannon's capacity equation as follows:

$$\frac{L}{\tau W_N} = \log_2(1 + \mu_r). \quad (5)$$

It should be noted that the effect of finite block length can be easily incorporated in the above expression by means of an SNR gap. Interested readers can refer to [26] for more details. Using (3) and (5), we can find the minimum transmission time required to transmit L bits over bandwidth W_N under the maximum transmit power P_{\max} constraint as:

$$\tau \geq \frac{L}{W_N \log_2(1 + P_{\max} \mu_N g)} = \tau_{\min}. \quad (6)$$

Similarly, the minimum transmission bandwidth required for the transmission of L bits over time τ is given by the solution of the following equation:

$$\frac{L}{\tau W_{\min}} = \log_2 \left(1 + P_{\max} \mu \left(\frac{W}{W_{\min}} \right) g \right). \quad (7)$$

As will be evident later, (6) and (7) will be useful in formulating the optimization problems for TDMA and FDMA cases, respectively. These arguments easily extend to the CDMA case and are discussed as a part of the uncoordinated strategies in the next section.

III. UNCOORDINATED TRANSMISSION

In uncoordinated transmission, we assume that L bits are transmitted in each transaction with the understanding that L would be small (≈ 50 bits) when only control information, such as device identity, is transmitted and would be relatively large (≈ 1000 bits) when both control information and data are transmitted together. For device multiplexing, we consider two strategies: i) CDMA random access, and ii) FDMA random access. We start this discussion with the design of CDMA random access, from which the FDMA random access design will follow.

A. CDMA Random Access

In CDMA random access, we assume that each device selects a N_c length code randomly from the set of $2^{N_c} - 1$ possible binary sequences, where N_c is the design parameter. As a result of this random code selection, the chosen codes will not be perfectly orthogonal in general. Since the transmissions of various devices are synchronized, the cross correlation in two randomly chosen codes is assumed to be $1/N_c$. For fair comparison across all the transmission strategies considered in this paper, we assume that the total bandwidth is W over which the CDMA waveform is transmitted, leading to W/N_c as the effective bandwidth of the information symbols. We further assume that the devices perform uplink power control such that the target SINR at the base station is μ_t .

1) *System Design*: As is clear from the setup, the only information devices have about the number of transmission requests $N_a \sim \text{Pois}(\lambda \tau_s)$ is the mean load $\lambda \tau_s$. Therefore, we design the system for $(1-\epsilon)^{th}$ percentile of the arrivals, which is denoted by \bar{N} , where $\mathbb{P}[N_a > \bar{N}] \leq \epsilon$. This also defines first failure event, i.e., when the actual number of arrivals are greater than \bar{N} and it is no longer possible to achieve the target SINR while satisfying the maximum power constraint. We will use this later in our discussion to bound the overall failure probability.

After determining \bar{N} from the load information, the next step is to determine the length N_c of the spreading codes. For this, we first define the collision event as follows.

Definition 1 (Collision Event \mathcal{A}_c). *Collision is said to occur when more than one device choose the same spreading code.*

Clearly, collision leads to packet failures since there is no way to differentiate between various devices. For given \bar{N} and N_c , the probability of collision is given by the following Lemma.

Lemma 1. *Defining $m (= 2^{N_c} - 1)$ as the number of codes and \bar{N} as the number of arrivals, the probability of collision can be expressed as*

$$P_c = \mathbb{P}[\mathcal{A}_c] = 1 - \left(1 - \frac{1}{m}\right)^{\bar{N}-1}. \quad (8)$$

Proof: First note that the probability of a particular code being chosen by n devices is

$$\mathbb{P}[n] = \binom{\bar{N}}{n} \left(\frac{1}{m}\right)^n \left(1 - \frac{1}{m}\right)^{\bar{N}-n}. \quad (9)$$

The result follows from the fact that

$$P_c = 1 - \mathbb{P}[0] - \mathbb{P}[1] = \mathbb{P}[\mathcal{A}_c] \quad (10)$$

$$= 1 - \left(1 - \frac{1}{m}\right)^{\bar{N}} - \frac{1}{m} \left(1 - \frac{1}{m}\right)^{\bar{N}-1}. \quad (11)$$

Corollary 1. *The length of the spreading sequence required to restrict the collision probability within a predefined value P_c is*

$$N_c = \left\lceil \log_2 \left(1 + \frac{1}{1 - (1 - P_c)^{\frac{1}{\bar{N}-1}}} \right) \right\rceil \quad (12)$$

$$\approx \left\lceil \log_2 \left(1 + \frac{\bar{N} - 1}{P_c} \right) \right\rceil. \quad (13)$$

Proof: The main result follows directly from Lemma 1 and the approximation follows in the low collision probability regime from the fact that

$$\lim_{P_c \rightarrow 0} (1 - P_c)^{\frac{1}{\bar{N}-1}} \stackrel{(a)}{=} 1 + \frac{\ln(1 - P_c)}{\bar{N} - 1} \quad (14)$$

$$\stackrel{(b)}{=} 1 - \frac{P_c}{\bar{N} - 1}, \quad (15)$$

where (a) follows from the Taylor series expansion of a^x and (b) follows from the fact that $\lim_{x \rightarrow 0} \ln(1 - x) = -x$. An alternate way to directly prove the approximate result is by observing that the collision probability P_c can be tightly upper bounded by

$$P_c \leq \frac{\bar{N} - 1}{m}. \quad (16)$$

To prove this bound, define M as the random variable denoting the number of unique codes occupied by $\bar{N} - 1$ devices. The collision probability can now be expressed as

$$P_c = \frac{E[M]}{m} \leq \frac{\bar{N} - 1}{m}, \quad (17)$$

which completes the proof. ■

For a given N_c and channel gain g , the transmit power required to achieve target SINR μ_t can be derived in the closed form and is given by the following Lemma. This will be useful when we compare and contrast various access techniques in Section V.

Lemma 2 (CDMA Transmit Power). *The transmit power and the energy per bit required by a mobile to achieve target SINR $\mu_t = 2^{\frac{LN_c}{W\tau_s}} - 1$ with channel gain g and code length N_c are*

$$P_t = [\mu g (N_c \mu_t^{-1} - (\bar{N} - 1))]^{-1} \quad (18)$$

$$E_b = \frac{\tau_s}{L} [\mu g (N_c \mu_t^{-1} - (\bar{N} - 1))]^{-1}. \quad (19)$$

Proof: From Shannon's capacity equation, we have

$$\frac{LN_c}{W\tau_s} = \log_2(1 + \mu_t) \Rightarrow \mu_t = 2^{\frac{LN_c}{W\tau_s}} - 1 \quad (20)$$

Assuming cross correlation in spreading codes to be $1/N_c$ and the effective symbol bandwidth to be W/N_c , the target SINR μ_t can be expressed as

$$\mu_t = \frac{N_c}{(\bar{N} - 1) + \mu_f^{-1}} \Rightarrow \mu_f = [N_c \mu_t^{-1} - (\bar{N} - 1)]^{-1}, \quad (21)$$

where μ_f is the fixed received SNR at the base station. The required transmit power can now be expressed as

$$\mu_f = P_t \mu g \Rightarrow P_t = [\mu g (N_c \mu_t^{-1} - (\bar{N} - 1))]^{-1}, \quad (22)$$

which completes the proof. ■

2) *Failure Probability:* There are three failure events possible: i) $N_a > \bar{N}$, in which case we have $\mu_r < \mu_t$ and hence all the packets are in error, ii) collision event \mathcal{A}_c in which two users pick the same random sequence, and iii) P_t derived in Lemma 2 is more than P_{\max} for a particular device so that the maximum power constraint is violated (device is in outage). Denote this outage probability by δ . As a matter of convention, we assume that the devices in outage transmit at their maximum power. This is justified because it is highly likely (with probability $1 - \epsilon$) that the actual number of arrivals are much less than \bar{N} for which the system is designed. This may ensure successful transmission in certain cases even though the transmit power was less than the designed value. In this case, the outage probability δ defined as $\mathbb{P}[P_t > P_{\max}]$ is an upper bound on the actual outage. For the given failure probabilities ϵ , P_c and δ , the overall failure probability P_f is upper bounded by the following Proposition.

Proposition 1. *The overall failure probability in CDMA random access case is*

$$P_f \leq \epsilon + (1 - \epsilon)\{\delta + P_c(1 - \delta)\}, \quad (23)$$

where the inequality reduces to equality when $\delta = 0$.

Proof: For a general failure event \mathcal{A} , the failure probability $P_f = \mathbb{P}[\mathcal{A}]$ can be expressed as:

$$P_f = \mathbb{P}[\mathcal{A}|N_a > \bar{N}]\mathbb{P}[N_a > \bar{N}] + \mathbb{P}[\mathcal{A}|N_a \leq \bar{N}]\mathbb{P}[N_a \leq \bar{N}] \quad (24)$$

$$= \epsilon + (1 - \epsilon)\{\mathbb{P}[\mathcal{A}|N_a \leq \bar{N}, P_t > P_{\max}]\mathbb{P}[P_t > P_{\max}] + \mathbb{P}[\mathcal{A}|N_a \leq \bar{N}, P_t \leq P_{\max}]\mathbb{P}[P_t \leq P_{\max}]\}, \quad (25)$$

from which the result follows by observing that the failure events corresponding to the first and second probability terms are, respectively, device outage and collision of the chosen random sequences. ■

This result shows that given the overall failure probability, the error probabilities ϵ , P_c and δ can not be independently chosen. Observe that the above upper bound is tight for small values of δ , which leads to the following characterization of the maximum load supported by a base station.

Corollary 2 (Maximum Load). *The maximum arrival rate (load) supported by the base station for given failure probability p and given outage probability δ is*

$$\lambda_{\max} = \max \{ \lambda : \mathbb{P}(P_t(\lambda) > P_{\max}) \leq \delta, P_f \leq p \}. \quad (26)$$

Example 1 (CDMA Maximum Load). *For general system parameters $L = 1000$ bits, $W = 1$ MHz, $\mu = -3$ dB, $\tau_s = 1$ sec, and CDMA specific parameters $P_f = .05$, $\epsilon = .01$, $\delta = 0$, and no fading, the maximum load that a base station can handle under CDMA random access is $\lambda \approx 1350$ arrivals per second.*

B. FDMA Random Access

FDMA random access design follows on the same lines as that of CDMA random access, with the only difference that the devices now choose one of the N_f orthogonal channels and when the two devices choose different channels, there is

no interference. Using the collision probability result derived in Lemma 1, the number of orthogonal channels required in this case can be expressed as:

$$N_f = \frac{1}{1 - (1 - P_c)^{\frac{1}{N-1}}}, \quad (27)$$

from which the transmit power can be expressed as:

$$P_t = \frac{2^{\frac{LN_f}{W\tau_s}} - 1}{\mu g N_f}. \quad (28)$$

For this new transmit power expression, the maximum load supported by a base station can be defined in the same way as done for the CDMA case.

Example 2 (FDMA Maximum Load). *For the same system parameters as that of Example 1, the maximum load that a base station can handle under FDMA random access is $\lambda \approx 160$ arrivals per second, which is order of magnitude lower than that of CDMA case.*

IV. COORDINATED TRANSMISSIONS

In this section, we discuss the design of coordinated strategies. We assume that the dedicated uplink connections are already established in the random access stage as discussed in the previous section. As a result, the base station has complete information about the N_a arrivals, which are to be coordinated. As is clear from (6) and (7) derived in Section II, there is a minimum transmission time τ_{\min} and minimum bandwidth W_{\min} required for the successful transmission of a given payload L , which appears due to the constraint on maximum power at which a device can transmit. Due to these constraints on the minimum resources required, there is clearly a fundamental limit on the number of packets that can be scheduled in a given time-frequency resource block. This limit defines the maximum load a base station can handle and is dependent upon the resource partitioning strategy being employed. Since the devices farther out near the cell edge or in deep fade require more resources than the others, we assume that the base station deliberately drops a small fraction of these arrivals to increase the maximum load it can handle and to reduce the overall average transmit power or energy [27].

For concreteness, we assume that the base station always drops δ_1 fraction of the arrivals, while ensuring that the total system outage is always less than δ . We will comment on the relationship between δ_1 and δ later in this section. Denote the number of devices actually served in each slot by $K \leq (1 - \delta_1)N_a$. It should be noted that the dropped arrivals are the ones having smallest channel gains. For resource partitioning, we consider two approaches: i) TDMA – splitting time slot τ_s into K parts, and ii) FDMA – splitting bandwidth W into K parts, where the goal in both the approaches is to minimize total transmit power or energy. As shown later in this section, the optimization problems to find power and energy optimal schedules are convex and hence can be solved efficiently using known algorithms, such as the MoveRight algorithm [24]. However, we show that the numerical optimization is not necessary since a near-optimal tractable solution can be found where the resources allocated to each device solely depend

upon its channel gain and are independent of the channel gains of other devices. More interestingly, we analytically show that both the average transmit power and energy under equal resource allocation is within a small factor of the optimal values in both TDMA and FDMA. Furthermore, we also compare the results with the power optimal multiuser detection strategy, i.e., SIC, and show that FDMA (including simple equal bandwidth allocation) is power optimal in the limit of asymptotically low spectral efficiency.

In all the coordinated strategies discussed above, once the base station decides the resource allocation it is relayed to the devices as a part of the beacon signal, which is then used by the devices to transmit over the allocated resources. We first derive the power required in SIC, which is global optimal in terms of total power minimization and will serve as a benchmark for the orthogonal transmission strategies discussed later in this section.

A. Power Optimal Coordinated Strategy

For K users, assuming the channel gains are ordered in the increasing order, i.e., $g_1 \leq g_2 \leq \dots g_K$, the optimal strategy is the weakest-last interference cancellation strategy, in which the user with the strongest channel gain is decoded first assuming interference from the other users as noise [28]. This signal is then cancelled from the received signal while decoding the user with the second best channel gain. The process is repeated until the last user is decoded. Therefore, the user with the strongest channel gain sees interference from all the other users, whereas the user with the weakest channel gain does not see any interference. The total transmit power under this strategy is given by the following theorem.

Theorem 1. *For K users with $g_1 \leq g_2 \leq \dots g_K$, the total transmit power for the weakest-last successive interference cancellation strategy is:*

$$P = \frac{2^{\frac{L}{W\tau_s}} - 1}{\mu} \sum_{k=1}^K \frac{2^{\frac{(k-1)L}{W\tau_s}}}{g_k}. \quad (29)$$

Proof: Under this particular decoding strategy, the Shannon's capacity expression of the k^{th} user is:

$$\frac{L}{W\tau_s} = \log_2 \left(1 + \frac{P_k \mu g_k}{1 + \sum_{i=1}^{k-1} P_i \mu g_i} \right), \quad (30)$$

from which we can derive the following relationship in the transmit powers:

$$P_k \mu g_k = \left(2^{\frac{L}{W\tau_s}} - 1 \right) \left(1 + \sum_{i=1}^{k-1} P_i \mu g_i \right). \quad (31)$$

Starting from $k = 1$ and solving for the transmit powers P_k by substituting all the preceding values in the above expression leads to the following closed form expression for P_k :

$$P_k = \frac{2^{\frac{(k-1)L}{W\tau_s}} \left(2^{\frac{L}{W\tau_s}} - 1 \right)}{g_k \mu}, \quad (32)$$

which completes the proof. ■

We now discuss the TDMA system design in detail and then show that the results for FDMA case directly follow.

B. TDMA System Design

We first find the optimal power or energy partition of $\tau_s = \{\tau_1, \tau_2, \dots, \tau_K\}$ such that the i^{th} device transmits its L information bits over bandwidth W in time τ_i . Using Shannon's capacity expression, the total transmit power required under this allocation is:

$$P = \sum_{i=1}^K \frac{2^{\frac{L}{W\tau_i}} - 1}{\mu g_i}. \quad (33)$$

Similarly the total energy per bit can be expressed as:

$$E_b = \sum_{i=1}^K \frac{\tau_i}{L} \frac{(2^{\frac{L}{W\tau_i}} - 1)}{\mu g_i}. \quad (34)$$

As discussed above, the maximum power constraints leads to the following constraint on the transmission time:

$$\tau_i \geq \frac{L}{W \log_2(1 + P_{\max} \mu g_i)}. \quad (35)$$

The optimization problem to find the power (or energy) optimal schedule can now be formulated as follows:

$$\begin{aligned} \min_{\{\tau_i\}} \quad & \sum_{i=1}^K u(\tau_i) \\ \text{s.t.} \quad & \sum_{i=1}^K \tau_i \leq \tau_s \\ & \tau_i \geq \frac{L}{W \log_2(1 + P_{\max} \mu g_i)} (= \tau_{\min_i}) \\ & 1 \leq i \leq K, \end{aligned}$$

where the function $u(\tau_i)$ is chosen appropriately depending upon whether the goal is to minimize total power or energy. Regardless of this choice, the optimization problem remains convex and hence can be efficiently solved using known algorithms [24]. Clearly the power and energy optimal allocations are in general different for TDMA.

Remark 1 (Feasibility and Maximum Load). *The above optimization problem is feasible if the constraint $\sum_{i=1}^K \tau_{\min_i} \leq \tau_s$ is satisfied. Using this constraint, an approximation on the maximum load can be derived as follows:*

$$\frac{\sum_{i=1}^K \tau_{\min_i}}{K} \leq \frac{\tau_s}{K} \Rightarrow K \leq \frac{\tau_s}{\sum_{i=1}^K \tau_{\min_i} / K} \stackrel{(a)}{\approx} \frac{\tau_s}{\mathbb{E}[\tau_{\min}]}, \quad (36)$$

where (a) follows from two approximations: i) Strong Law of Large Numbers (SLLN) holds and the average converges to the mean of the random variables, ii) the mean is $\mathbb{E}[\tau_{\min}]$, which is not exact because δN_a arrivals with smallest channel gains have been removed and hence order statistics should be used to compute the mean of the remaining random variables. Nevertheless, this approximation is tight when $\delta \rightarrow 0$, which is the regime of interest in this system design. More formally, the maximum load can be defined in terms of a given outage constraint as follows:

$$\lambda_{\max} = \max \{ \lambda : \mathbb{P}[N_a(1 - \delta_1) \geq K_{\max}] \leq \epsilon_1 \}, \quad (37)$$

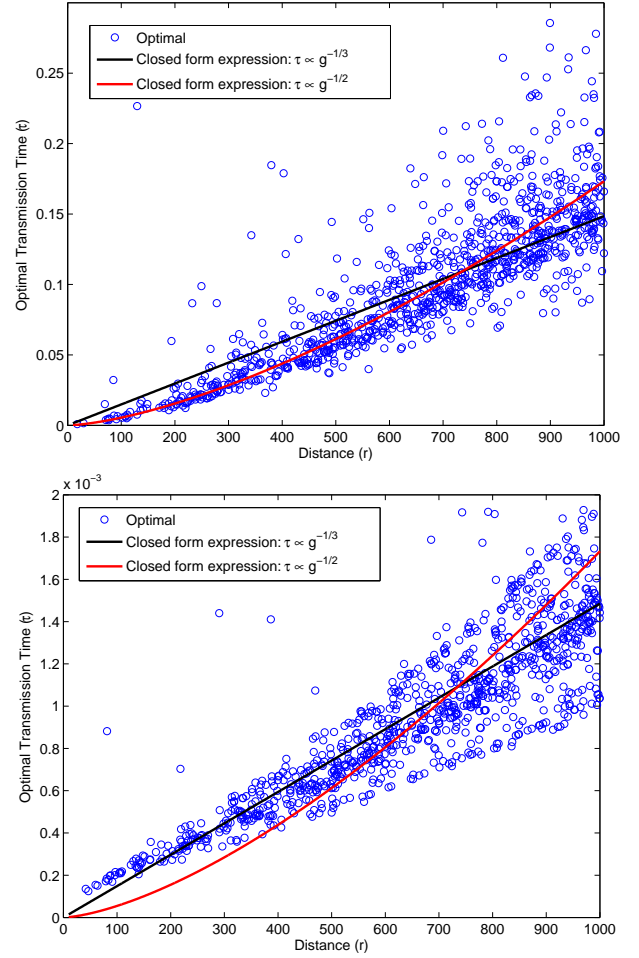


Fig. 4. Comparison of the closed form solution given by (47) with the power optimal solution (scatter plot). (first) low load ($\lambda = 10$). (second) high load ($\lambda = 1000$). The fading and shadowing is averaged out.

where K_{\max} is defined by ordering the arrivals in terms of decreasing channel gains as follows:

$$K_{\max} = \max_K \sum_{i=1}^K \tau_i \leq \tau_s. \quad (38)$$

The parameters δ_1 and ϵ_1 should be chosen such that the total system outage is less than δ . This is made precise in the following Proposition.

Proposition 2 (TDMA Outage). *The total system outage δ can be upper bounded in terms of the failure probabilities δ_1 and ϵ_1 as*

$$\delta \leq \epsilon_1 + \delta_1(1 - \epsilon_1). \quad (39)$$

Proof: Using similar ideas as in Proposition 1, the total outage δ can be written in terms of the failure event \mathcal{A} as:

$$\delta = \mathbb{P}[\mathcal{A} | (1 - \delta_1)N_a \geq K_{\max}] \mathbb{P}[(1 - \delta_1)N_a \geq K_{\max}] + \mathbb{P}[\mathcal{A} | (1 - \delta_1)N_a < K_{\max}] \mathbb{P}[(1 - \delta_1)N_a < K_{\max}] \quad (40)$$

$$= \epsilon_1 + \delta_1(1 - \epsilon_1), \quad (41)$$

where we bounded the term $\mathbb{P}[\mathcal{A} | (1 - \delta_1)N_a \geq K_{\max}]$ by 1. ■

This bound is tight especially when $\epsilon_1 \rightarrow 0$, which will be the regime of interest for this paper.

1) *Near Optimal Closed-Form Solution:* The form of the optimization problem is such that the exact closed form solutions are not possible. We now show that it is possible to derive approximate near-optimal closed form results, where the transmission time is solely a function of the channel gain of the device of interest independent of the channel gains of other devices. We demonstrate it for the power minimization problem, where the total transmit power can be expressed as:

$$P = \sum_{i=1}^K \frac{2^{\frac{L}{W\tau_i}} - 1}{\mu g_i} \quad (42)$$

$$= \frac{2^{\frac{L}{W\tau_1}} - 1}{\mu g_1} + \frac{2^{\frac{L}{W\tau_2}} - 1}{\mu g_2} \dots \frac{2^{\frac{L}{W\left(\tau_s - \sum_{j=1}^{K-1} \tau_j\right)} - 1}}{\mu g_K}. \quad (43)$$

Minimizing the transmit power P w.r.t. τ_1 we get

$$\frac{\delta P}{\delta \tau_1} = 0 \Rightarrow \frac{2^{\frac{L}{W\tau_1}}}{g_1 \tau_1^2} = \frac{2^{\frac{L}{W\tau_N}}}{g_N \tau_N^2} \Rightarrow \tau_i^2 2^{\frac{L}{W\tau_i}} \propto g_i^{-1} \quad (44)$$

Remark 2. For small λ , τ is of the order of seconds, and for our choice of L and W , $L/W = 10^{-3}$, which implies $2^{\frac{L}{W\tau}} \rightarrow 1$. Therefore, $\tau \propto g^{-1/2}$. On the other hand, for large λ (say $\lambda = 1000$), and $L/W = 10^{-3}$, we have $2^{\frac{L}{W\tau}} \approx \frac{2L}{W\tau}$. Therefore, $\tau \propto g^{-1/3}$ in this regime.

From the above remark, we note that the transmission time can be expressed solely as a function of channel gain as

$$\tau_i = \frac{f(g_i)}{\sum_j f(g_j)} \tau_s, \quad (45)$$

which reduces further to

$$\tau_i = \frac{f(g_i)}{\lambda E[f(g)]}, \quad (46)$$

for reasonably high values of λ . Ignoring fading and shadowing and assuming $f(g_i) = g_i^{-1/n}$ that encompasses both the special cases of $n = 2$ and $n = 3$, it can be expressed in the closed form as follows:

$$\tau_i = \frac{2 + \gamma/n}{2\lambda} r^{\gamma/n} \frac{r_0^2 - r_i^2}{r_0^{2+\gamma/n} - r_i^{2+\gamma/n}}. \quad (47)$$

In Fig. 4, we plot the closed form result (47) along with a scatter plot of the optimal transmission times in two regimes: i) low load ($\lambda = 10$) and ii) high load ($\lambda = 1000$). We note that the approximations for both the regimes are quite accurate, which is more surprising for the low load case since the closed form expression was derived assuming that the value of λ is reasonably high. Interested readers can refer to [1] for further details on the accuracy of this simple approach.

For the energy optimal solution, we present a similar comparison in Fig. 5, which shows that the approximation $\tau \propto g^{-1/3}$ is accurate both in the low and high load regimes. This leads to the same closed form solution in both the regimes.

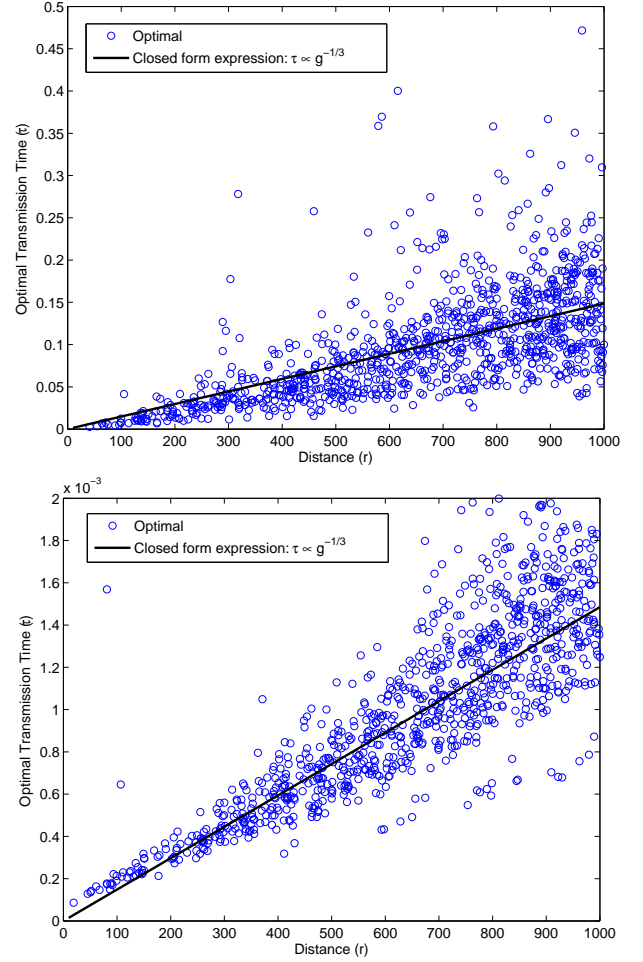


Fig. 5. Same as Fig.4 except that the scatter point now corresponds to energy optimal solution. (first) low load ($\lambda = 10$). (second) high load ($\lambda = 1000$). The same closed form solution works in both the regimes.

2) *Equal Time vs Power Optimal Allocation:* We now consider an even simpler case in which each device is allocated same transmission time and show that the transmit power under this simple strategy is always within a small constant of the optimal power in the parameter space of interest. The result is given by the following Theorem and the proof is given in Appendix A.

Theorem 2. The ratio of total transmit powers of K devices under uniform (U_{g_1, g_2, \dots, g_K}) and optimal (P_{g_1, g_2, \dots, g_K}) schedules is upper bounded by:

$$\frac{U_{g_1, g_2, \dots, g_K}}{P_{g_1, g_2, \dots, g_K}} \leq \frac{2^{\frac{KL}{W\tau_s}} - 1}{2^{\frac{L}{W\tau_s} \frac{\sum_{j=1}^K \sqrt{\frac{g_1}{g_j}}}{L}} - 1}, \quad (48)$$

where $g_1 = \min\{g_j\}$.

This bound is surprisingly tight as shown in the following example.

Example 3 (Bound in the parameter space of interest). For $K = 1000$, $\tau_s = 1$ sec, $L = 1000$ bits, $W = 1$ MHz and no fading, the bound given by (48) is ≈ 2 , i.e., it is guaranteed that the transmit power under equal resource allocation is

no more than around 3 dB worse than the optimal power. In [1] we have numerically shown that the actual gap is much smaller.

3) *Equal Time vs Energy Optimal Allocation:* On the same lines, we now compare the energy per bit required under equal time allocation with that of the energy optimal allocation. Before deriving the main result, we first derive an upper bound on the optimal time allocated to the device corresponding to the minimum channel gain. The proof is given in Appendix B.

Lemma 3. *The optimal time allocated to the device corresponding to the minimum channel gain can be upper bounded as*

$$\tau_1^* \leq \frac{\tau_s}{\sum_{i=1}^K g_1/g_i}, \quad (49)$$

where $g_1 = \min\{g_j\}$.

From this Lemma the following result follows. The proof is in Appendix C.

Theorem 3. *The ratio of the total energy per bit under uniform (U_{g_1, g_2, \dots, g_K}) and optimal (E_{g_1, g_2, \dots, g_K}) schedules is upper bounded by:*

$$\frac{U_{g_1, g_2, \dots, g_K}}{E_{g_1, g_2, \dots, g_K}} \leq \frac{\sum_{i=1}^K g_1/g_i}{K} \frac{2^{\frac{KL}{W\tau_s}} - 1}{2^{\frac{L \sum_{i=1}^K g_1/g_i}{W\tau_s}} - 1}, \quad (50)$$

where $g_1 = \min\{g_j\}$.

Corollary 3. *In the limit of asymptotically low spectral efficiency, equal resource allocation is sum energy optimal on the space of TDMA strategies, i.e.,*

$$\lim_{W\tau_s \rightarrow \infty} \frac{U_{g_1, g_2, \dots, g_K}}{E_{g_1, g_2, \dots, g_K}} = 1. \quad (51)$$

The proof of this corollary directly follows from Theorem 3 under the limit $\frac{L}{W\tau_s} \rightarrow 0$ by using $\lim_{x \rightarrow 0} a^x - 1 = x \ln(a)$. In addition to this asymptotic result, the bound is surprisingly tight even in the parameter space of interest for M2M. This is shown in the following example.

Example 4. *For the same system parameters as that of Example 3, the ratio of energy per bit in the uniform and optimal allocation strategies is bounded above by 1.25. Clearly, the value of optimization in terms of energy per bit is also limited for TDMA.*

C. FDMA System Design

The FDMA system design proceeds exactly in the same way as discussed for the TDMA case above. We will therefore highlight only the main differences in problem formulation. The goal here is to partition the available bandwidth W into K parts $\{W_1, W_2, \dots, W_K\}$ so as to minimize the total transmit power or energy. The total power required in this case can be expressed as:

$$P = \sum_{i=1}^K \frac{W_i}{W} \frac{2^{\frac{L}{W_i\tau_s}} - 1}{\mu g_i}, \quad (52)$$

and the total energy per bit can be expressed as:

$$E_b = \frac{\tau_s}{L} \sum_{i=1}^K \frac{W_i}{W} \frac{(2^{\frac{L}{W_i\tau_s}} - 1)}{\mu g_i}. \quad (53)$$

Unlike TDMA case, both the power and energy optimal schedules are exactly the same in case of FDMA. This is simply because the energy per bit can be expressed as a constant multiple of power, where the constant is independent of the optimization parameters. The constraint on the maximum transmit power translates to the minimum bandwidth required by each device depending upon its channel condition. This minimum bandwidth W_{\min} is the solution of the following equation:

$$\frac{L}{\tau_s W_{\min_i}} = \log_2 \left(1 + P_{\max} \mu \left(\frac{W}{W_{\min_i}} \right) g_i \right). \quad (54)$$

The energy or power minimization problem can now be formulated as:

$$\begin{aligned} \min_{\{W_i\}} \quad & \sum_{i=1}^K v(W_i) \\ \text{s.t.} \quad & \sum_{i=1}^K W_i \leq W \\ & W_i \geq W_{\min_i} \\ & 1 \leq i \leq K, \end{aligned}$$

where $v(W_i)$ is the cost function representing power or energy per bit. We note that the form of the optimization problem is exactly the same as that of the TDMA problem discussed above in detail and hence most of the insights about the exact and approximate solutions carry over.

Remark 3 (Feasibility and Maximum Load). *As discussed for the TDMA counterpart in Remark 1, the optimization problem is feasible if the constraint $\sum_{i=1}^K W_{\min_i} \leq W$ is satisfied. Using this constraint, an approximate bound on K can be derived as:*

$$K \leq \frac{W}{\mathbb{E}[W_{\min}]}. \quad (55)$$

The maximum load can also be defined in the same way as done for the TDMA case.

This completes the analysis of the coordinated strategies and we now compare the maximum load that can be handled by a base station using TDMA and FDMA in the following example.

Example 5 (Maximum Load: FDMA vs TDMA). *Choosing the same set of general system parameters as that of Example 1 and δ vanishingly small, the maximum load a base station can handle using TDMA and FDMA is ≈ 1200 and ≈ 14700 respectively. This clearly shows that it is optimal to partition over frequency. Although TDMA and FDMA are exactly the same from information theoretic sense, the difference in the optimal solution arises as a result of the peak power constraint that affects the two schemes differently as is apparent from the expressions of τ_{\min} and W_{\min} . We will comment more on this in Section V.*

As done for the TDMA case, we now compare equal resource allocation with the optimal allocation in terms of transmit power and energy per bit.

1) *Equal Bandwidth Allocation vs Optimal Allocation:*

Before deriving the main result, as in the TDMA case we first derive an upper bound on the optimal bandwidth allocated to the packet corresponding to the minimum channel gain. The proof follows on the same lines as that of Lemma 3 and a sketch is given in Appendix D for completeness.

Lemma 4. *The optimal bandwidth allocated to the packet corresponding to the minimum channel gain can be upper bounded by*

$$W_1^* \leq \frac{W}{\sum_{i=1}^K g_1/g_i}, \quad (56)$$

where $g_1 = \min\{g_j\}$.

From this Lemma, the following bound on the ratios of the total transmit powers follows. The same bound holds for the total energy per bit as well. The proof follows on the same lines as that of Theorem 3 and is hence skipped.

Theorem 4. *The ratio of total transmit powers (and energies) under uniform and optimal schedules can be bounded as:*

$$\frac{U_{g_1, g_2 \dots g_K}}{P_{g_1, g_2 \dots g_K}} \leq \frac{\sum_{i=1}^K g_1/g_i}{K} \frac{2^{\frac{KL}{W\tau_s}} - 1}{2^{\frac{L \sum_{i=1}^K g_1/g_i}{W\tau_s}} - 1}, \quad (57)$$

where $g_1 = \min\{g_j\}$.

Remark 4 (FDMA Bound and Optimality). *Note that this bound is the same as the one derived for the energy optimal solution of TDMA. Therefore, Example 4 is applicable in this case and hence the ratio of energy or power under uniform and optimal allocation strategies is upper bounded by 1.25. Moreover, equal resource allocation is both sum power and sum energy optimal over the space of FDMA strategies. This follows from Corollary 3.*

2) *Equal Bandwidth Allocation vs SIC:* We conclude this section with an even stronger result, which proves that FDMA is sum power optimal over the space of general resource allocation strategies – not limited to orthogonal – in the limit of low spectral efficiency. The result follows by comparing the transmit power under equal bandwidth allocation with the global optimal transmit power achieved by SIC and is given by the following theorem. We denote the transmit power under equal bandwidth allocation by $U_{g_1, g_2 \dots g_K}$ and by slight abuse of notation under SIC by $P_{g_1, g_2 \dots g_K}$.

Theorem 5. *Equal bandwidth allocation is sum power optimal over a space of general resource allocation strategies in the limit of low spectral efficiency, i.e.,*

$$\lim_{W\tau_s \rightarrow \infty} \frac{U_{g_1, g_2 \dots g_K}}{P_{g_1, g_2 \dots g_K}} = 1. \quad (58)$$

Proof: From the expressions of the sum power under SIC given by (29) and under FDMA given by (52), the ratio can

be expressed as

$$\frac{U_{g_1, g_2 \dots g_K}}{P_{g_1, g_2 \dots g_K}} = \frac{1}{K} \frac{2^{\frac{LK}{W\tau_s}} - 1}{2^{\frac{L}{W\tau_s}} - 1} \frac{\sum_{k=1}^K \frac{1}{g_k}}{\sum_{k=1}^K \frac{1}{g_k} 2^{\frac{(k-1)L}{W\tau_s}}}, \quad (59)$$

from which the result follows by using $\lim_{x \rightarrow 0} a^x - 1 = x \ln(a)$. ■

V. DISCUSSION AND NUMERICAL RESULTS

Note that the preferred choices for both the uncoordinated and coordinated strategies are clear from examples 2 and 5. In case of uncoordinated strategies, example 2 shows that the random access CDMA supports order of magnitude higher load than random access FDMA and in case of coordinated strategies, example 5 shows that coordinated FDMA is clearly a better choice over coordinated TDMA. Therefore, in this section we will not consider random access FDMA and coordinated TDMA, except when we compare energy optimal solutions. Interested readers can refer to [1] for a detailed discussion on the optimal power TDMA results. Instead, our main focus in this section will be on random access CDMA, coordinated FDMA and SIC in various regimes of interest.

A. Optimal Transmit Power

We first compare the optimal transmit power required for random access CDMA, coordinated FDMA and SIC in Fig. 6, where we plot both the mean power and the 95th percentile of the power. In both the cases, we observe that the performance of random access CDMA is close to that of FDMA in low and moderate arrival rates, especially if one accounts for the signaling overhead required in FDMA case. This shows that CDMA random access may be preferred at low to moderate arrival rates due to the apparent simplicity of the resulting system design. As expected, FDMA is the only option at high arrival rates and its performance is very close to that of the optimal SIC performance. One possible system design for this regime is to first use random access CDMA to establish uplink connection by sending small payload containing only the control information. The load supported by random access CDMA in this case will be very high because of the small payload size. The base station can then decide about the frequency allocation and relay this information back to the devices through broadcast signals. Recall that these two system designs, corresponding to high and low loading, are exactly the same as the one stage and two stage design examples discussed in Section II.

B. Optimal Energy per Bit

We now compare the three strategies along with coordinated TDMA in terms of the energy per bit in Fig. 7. We first note that the performance of TDMA and FDMA cases is similar. This is intuitive because both TDMA and FDMA are exactly the same from information theoretic sense. The difference in the transmit powers required in both the cases is a result of the peak power constraint that affects the two schemes differently. Now comparing the three main candidate strategies, we again note that random access CDMA performance is close to

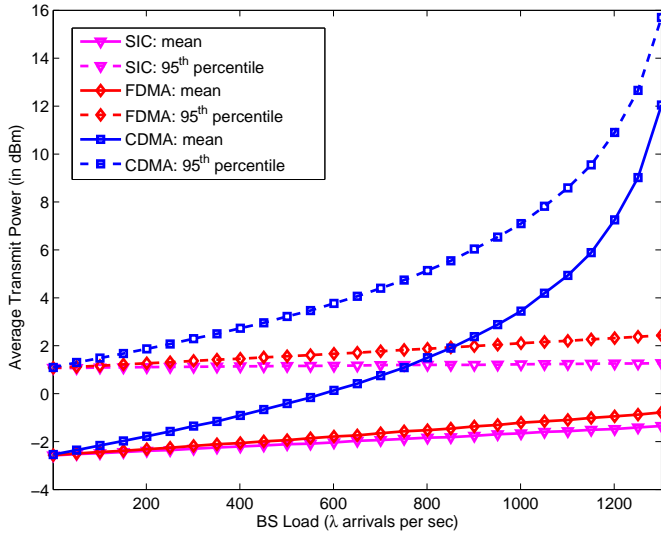


Fig. 6. Comparison of power optimal solutions of SIC and FDMA with random access CDMA.

that of coordinated FDMA at low to moderate loading, and coordinated FDMA is the only choice at high loading. The SIC performance is close to that of FDMA in this case as well. These observations lead to the same design guidelines as discussed above in case of optimal transmit power.

C. Coordinated FDMA vs SIC at very High Loading

In Fig. 8, we consider very high loading regime and restrict our comparison to SIC and coordinated FDMA for which we consider both optimal and equal bandwidth allocation strategies. We evaluate both the mean optimal power and 95th percentile of the power for both the strategies. Comparing equal bandwidth allocation with optimal FDMA, we note that equal resource allocation is near-optimal even at very high loading, which is consistent with our analysis. On the other hand, we observe comparatively higher gap between the two strategies in terms of 95th percentile of the power. The gap is, however, not too significant since even at $\lambda = 6000$ the transmit powers differ by less than 3 dB. On the other hand, the optimal SIC performance is comparatively much better, with the performance gap being more than 6 – 7 dB from the equal bandwidth allocation and around 4 dB from optimal bandwidth allocation at $\lambda = 6000$. However, it is important to note that optimal SIC requires perfect channel estimates for all the devices, which may not be realistic especially at high arrival rates. Therefore, the performance gap, even at very high loading, is unlikely to be significant if we account for high implementation losses expected in SIC. Exact quantification of these losses is out of the scope of this paper.

VI. CONCLUSIONS

In this paper, we have developed a systematic framework to study the power and energy optimal system design in the parameter space of interest for M2M communications. For comparison, we consider a variety of uncoordinated strategies, such as random access CDMA and FDMA, and coordinated

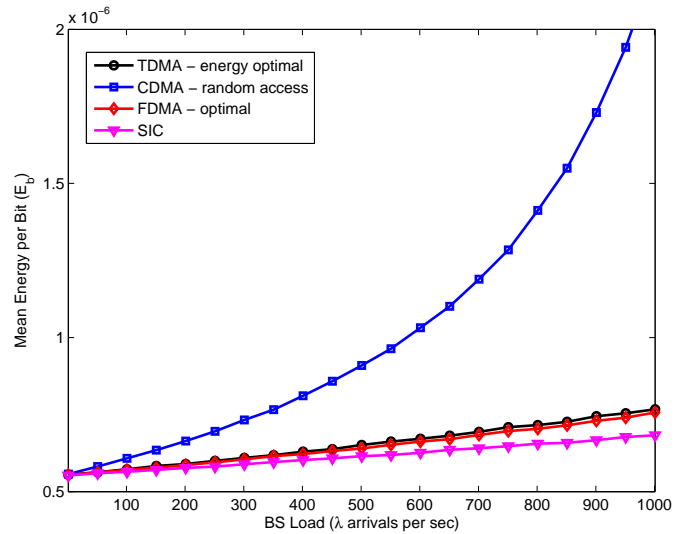


Fig. 7. Comparison of energy optimal solution of TDMA and FDMA with random access CDMA.

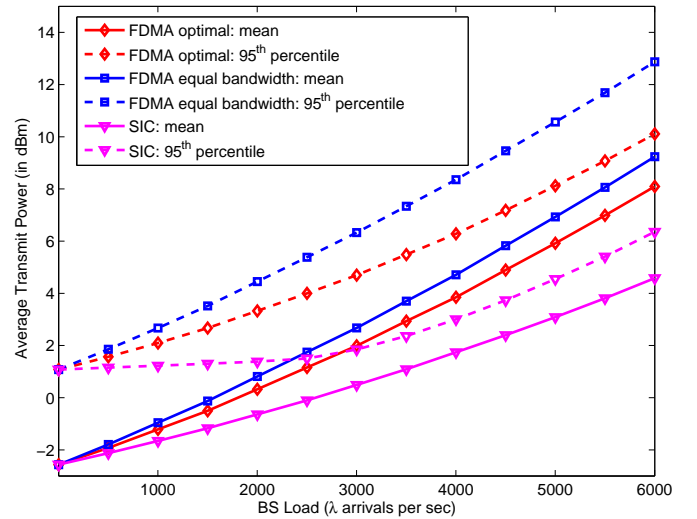


Fig. 8. Comparison of the optimal power and equal bandwidth allocation solutions for FDMA.

strategies, such as SIC, FDMA and TDMA. While the coordinated FDMA is the best practical strategy overall, random access CDMA is almost as good when the base station is lightly loaded and is a strong candidate for the actual system design due to its apparent simplicity and no signaling overhead. Additionally, we have shown that the value of optimization is in general small both for TDMA and FDMA, and simpler resource allocation strategies, such as equal resource allocation achieve performance within a small constant of the optimal performance. In the limit of asymptotically small spectral efficiency, equal bandwidth allocation in coordinated FDMA is shown to be power optimal over the space of general resource allocation strategies. An important extension of this work includes accounting for the implementation losses and signaling overhead while comparing various multiple access strategies.

APPENDIX A PROOF OF THEOREM 2

Let U_{g_1, g_2, \dots, g_K} denote the total transmit power under uniform schedule (equal transmission time) and P_{g_1, g_2, \dots, g_K} denote the power under the optimal schedule. Further let U_{g_i, \dots, g_j} and P_{g_i, \dots, g_j} be the transmit powers of the subset of users under uniform and optimal schedules, respectively. Without loss of generality, we assume that the channel gains g_i are indexed in the increasing orders of magnitude. Let the total transmission time of K devices be τ , with τ_j^* representing the optimal transmission time of j^{th} packet. The ratio of powers under two schemes can now be expressed as:

$$\frac{U_{g_1, g_2, \dots, g_K}}{P_{g_1, g_2, \dots, g_K}} = \frac{U_{g_1} + U_{g_2} + \dots + U_{g_K}}{P_{g_1} + P_{g_2} + \dots + P_{g_K}} \quad (60)$$

$$\leq \max \left\{ \frac{U_{g_1}}{P_{g_1}}, \frac{U_{g_2}}{P_{g_2}}, \dots, \frac{U_{g_K}}{P_{g_K}} \right\}. \quad (61)$$

The ratio U_{g_k}/P_{g_k} can be expressed as:

$$\frac{U_{g_k}}{P_{g_k}} = \frac{2^{\frac{KL}{W\tau_s}} - 1}{2^{\frac{L}{W\tau_k^*}} - 1}. \quad (62)$$

From the following optimality condition derived in (44):

$$\frac{2^{L/W\tau_k^*}}{g_k \tau_k^{*2}} = \frac{2^{L/W\tau_1^*}}{g_1 \tau_1^{*2}}, \quad (63)$$

we note that $\tau_1^* \geq \tau_k^* \forall k$ since $g_1 \leq g_k$. Therefore, (61) can be expressed as:

$$\frac{U_{g_1, g_2, \dots, g_K}}{P_{g_1, g_2, \dots, g_K}} \leq \frac{U_{g_1}}{P_{g_1}}. \quad (64)$$

To bound $\frac{U_{g_1}}{P_{g_1}}$, we first derive the following inequality from the optimality condition:

$$\tau_k^* \geq \tau_1^* \sqrt{\frac{g_1}{g_k}}, \forall k. \quad (65)$$

Using (65), we now derive an upper bound on τ_1^* as follows:

$$\tau_s = \sum_{j=1}^K \tau_j^* \geq \tau_1^* \sum_{j=1}^K \sqrt{\frac{g_1}{g_j}} \Rightarrow \tau_1^* \leq \tau_s / \sum_{j=1}^K \sqrt{\frac{g_1}{g_j}}. \quad (66)$$

Using (66), the ratio U_{g_1}/P_{g_1} can be bounded as:

$$\frac{U_{g_1}}{P_{g_1}} = \frac{2^{KL/W\tau_s} - 1}{2^{L/W\tau_1^*} - 1} \leq \frac{2^{\frac{KL}{W\tau_s}} - 1}{2^{\frac{L}{W\tau_s} \sum_{j=1}^K \sqrt{\frac{g_1}{g_j}}} - 1}, \quad (67)$$

which completes the proof. \blacksquare

APPENDIX B PROOF OF LEMMA 3

Assume that the channel gains g_i are indexed in the increasing orders of magnitude. For notational simplicity, denote the transmit energy per bit of the i^{th} device transmitting for time $0 < x \leq \tau_s$ by $\Psi_i(x)$:

$$\Psi_i(x) = \frac{x 2^{\frac{L}{xW}} - 1}{L \mu g_i}. \quad (68)$$

Note that $\Psi_i(x)$ is a monotonically decreasing function in x . Therefore,

$$\begin{aligned} \Psi_i'(x) &= \frac{\delta}{\delta x} \Psi_i(x) \\ &= \frac{1}{L \mu g_i} \left(2^{\frac{L}{xW}} - 1 - 2^{\frac{L}{xW}} \frac{L}{xW} \ln 2 \right) \leq 0, \quad x > 0. \end{aligned} \quad (69)$$

Now we derive the optimality condition for the energy optimal TDMA allocation. The total transmit energy per bit can be expressed as:

$$E = \sum_{i=1}^K \Psi_i(\tau_i) = \Psi_1(\tau_1) + \Psi_2(\tau_2) + \dots + \Psi_K \left(\tau_s - \sum_{i=1}^{K-1} \tau_i \right). \quad (71)$$

Minimizing the transmit energy E w.r.t. τ_1 we get

$$\frac{\delta E}{\delta \tau_1} = 0 \Rightarrow \Psi_1'(\tau_1) = \Psi_j'(\tau_j) \quad \forall j, \quad (72)$$

which has to be satisfied for the optimal transmission time τ_i^* as well. Therefore,

$$\frac{g_1}{g_i} = \frac{2^{\frac{L}{\tau_1^* W}} \frac{L}{\tau_1^* W} \ln 2 + 1 - 2^{\frac{L}{\tau_1^* W}}}{2^{\frac{L}{\tau_i^* W}} \frac{L}{\tau_i^* W} \ln 2 + 1 - 2^{\frac{L}{\tau_i^* W}}} \quad (73)$$

$$= \frac{\frac{L}{\tau_1^* W}}{\frac{L}{\tau_i^* W}} \left(\frac{2^{\frac{L}{\tau_1^* W}} \ln 2 - \frac{\tau_1^* W}{L} (2^{\frac{L}{\tau_1^* W}} - 1)}{2^{\frac{L}{\tau_i^* W}} \ln 2 - \frac{\tau_i^* W}{L} (2^{\frac{L}{\tau_i^* W}} - 1)} \right) \quad (74)$$

$$\stackrel{(a)}{\leq} \frac{\tau_i^*}{\tau_1^*}, \quad (75)$$

where (a) follows from (i) the function $\Phi(x) = 2^{\frac{1}{x}} \ln 2 - x(2^{\frac{1}{x}} - 1)$ is a decreasing function of x and hence the ratio $\Phi(x_1)/\Phi(x_2) \leq 1$ for $x_1 \geq x_2$, and (ii) $\tau_1^* \geq \tau_i^*, \forall i$ which follows from the optimality condition along with the fact that $\Psi_i'(x)$ is a monotonic function of x . Note that

$$\tau_s = \sum_i \tau_i^* \geq \tau_1^* \sum_{i=1}^K \frac{g_1}{g_i} \Rightarrow \tau_1^* \leq \frac{\tau_s}{\sum_{i=1}^K g_1/g_i}, \quad (76)$$

which completes the proof. \blacksquare

APPENDIX C PROOF OF THEOREM 3

Let U_{g_1, g_2, \dots, g_K} denote the total energy per bit under uniform schedule (equal transmission time) and E_{g_1, g_2, \dots, g_K} denote the energy per bit under the optimal schedule. Further let U_{g_i, \dots, g_j} and P_{g_i, \dots, g_j} be the transmit powers of the subset of users under uniform and optimal schedules, respectively. Rest of the setup remains the same as that of the proof of Theorem 2. The ratio of energies under two scheduling schemes can now be expressed as:

$$\frac{U_{g_1, g_2, \dots, g_K}}{E_{g_1, g_2, \dots, g_K}} = \frac{U_{g_1} + U_{g_2} + \dots + U_{g_K}}{E_{g_1} + E_{g_2} + \dots + E_{g_K}} \quad (77)$$

$$\leq \max \left\{ \frac{U_{g_1}}{E_{g_1}}, \frac{U_{g_2}}{E_{g_2}}, \dots, \frac{U_{g_K}}{E_{g_K}} \right\}. \quad (78)$$

The ratio U_{g_k}/E_{g_k} can be expressed as:

$$\frac{U_{g_k}}{E_{g_k}} = \frac{\tau_s}{K\tau_k^*} \frac{2^{\frac{KL}{W\tau_s}} - 1}{2^{\frac{L}{\tau_k^*W}} - 1}. \quad (79)$$

From the optimality condition derived in Lemma 3, $\tau_1^* \geq \tau_i^*$ $\forall i$ since $g_1 \leq g_k$. Therefore, (78) can be expressed as:

$$\frac{U_{g_1, g_2, \dots, g_K}}{E_{g_1, g_2, \dots, g_K}} \leq \frac{U_{g_1}}{E_{g_1}}. \quad (80)$$

As stated in the proof of Lemma 3, E_{g_1} is a monotonically decreasing function of the transmission time. Hence the upper bound on the ratio follows from the result of Lemma 3. ■

APPENDIX D PROOF OF LEMMA 4

Denote the transmit powers of the subset of the users under equal bandwidth allocation and optimal allocation by U_{g_i, \dots, g_j} and P_{g_i, \dots, g_j} , respectively. Without loss of generality, assume that the channel gains g_i are indexed in the increasing orders of magnitude. Let the total bandwidth to be allocated to K packets is W , with W_j^* representing the optimal bandwidth of j^{th} packet. For notational simplicity, denote the transmit power of the i^{th} user using bandwidth $0 < x < W$ by $\Psi_i(x)$:

$$\Psi_i(x) = \left(\frac{x}{W}\right) \frac{2^{\frac{x}{x\tau}} - 1}{\mu g_i}. \quad (81)$$

Note that the functional form of Ψ_i is the same as that in Lemma 3. Therefore, the proof essentially follows on the same lines. Expressing the total transmit power as:

$$P = \sum_{i=1}^K \Psi_i(W_i) \quad (82)$$

$$= \Psi_1(W_1) + \Psi_2(W_2) + \dots + \Psi_K\left(W - \sum_{i=1}^{K-1} W_i\right), \quad (83)$$

and following the same methodology as of Lemma 3, we can derive the following lower bound on the ratio of channel gains:

$$\frac{g_1}{g_i} \leq \frac{W_i^*}{W_1^*}. \quad (84)$$

Now note that

$$W = \sum_i W_i^* \geq W_1^* \sum_{i=1}^K \frac{g_1}{g_i} \Rightarrow W_1^* \leq \frac{W}{\sum_{i=1}^K g_1/g_i}, \quad (85)$$

which completes the proof. ■

REFERENCES

- [1] H. S. Dhillon, H. C. Huang, H. Viswanathan, and R. A. Valenzuela, "On resource allocation for machine-to-machine (M2M) communications in cellular networks," in *Proc., IEEE Globecom Workshops*, Anaheim, CA, Dec. 2012.
- [2] The FocalPoint Group, "M2M white paper: the growth of device connectivity," white paper, 2003, available online: goo.gl/S0j5O.
- [3] Vodafone, "Global machine to machine communication," white paper, 2010, available online: goo.gl/4V8az.
- [4] Ericsson, "Device connectivity unlocks value," white paper, Jan. 2011, available online: goo.gl/alou6.
- [5] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 66 – 74, Apr. 2011.
- [6] GSMA, "Experience a world where everything intelligently connects: The Connected Life," white paper, Feb. 2012, available online: goo.gl/gaqeD.
- [7] Ericsson, "More than 50 billion connected devices," white paper, Feb. 2011, available online: goo.gl/vjY00.
- [8] H. Viswanathan, K. Mun, and F. Scholler, "Modeling and analysis of cellular wireless machine-to-machine communication traffic," submitted to *IEEE Journal on Sel. Areas in Communications*, Apr. 2012.
- [9] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "A first look at cellular machine-to-machine traffic – large scale measurement and characterization," London, Jun. 2012.
- [10] Y. Jou, R. Attar, S. Ray, J. Ma, and X. Zhang, "M2M over CDMA2000 1x case studies," in *Proc., IEEE Wireless Communications and Networking Conf. (WCNC)*, Cancun, Mexico, Mar. 2011, pp. 1546 – 1551.
- [11] *Health Informatics - PoC Medical Device Communication - Part 00101: Guide-Guidelines for the Use of RF Wireless Technology*, IEEE Std 11073-00101-2008, Dec. 2008.
- [12] A. Ghosh, J. Zhang, J. G. Andrews, and R. Muhamed, *Fundamentals of LTE*. Prentice-Hall, 2010.
- [13] *Study on RAN improvements for Machine-Type Communications*, 3GPP TR 37.868, 2010.
- [14] K. Zheng, F. Hu, W. Wang, W. Xiang, and M. Dohler, "Radio resource allocation in LTE-advanced cellular networks with M2M communications," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 184–192, Jul. 2012.
- [15] Y. Chen and W. Wang, "Machine-to-machine communication in LTE-A," in *Proc., IEEE Veh. Technology Conf. (VTC)*, Ottawa, Canada, Sep. 2010.
- [16] M. Martsola, T. Kiravuo, and J. K. O. Lindqvist, "Machine to machine communication in cellular networks," in *Proc., IEE Mobility Conference*, Guangzhou, China, Nov. 2005.
- [17] Y. Chen and Y. Yang, "Cellular based machine to machine communication with un-peer2peer protocol stack," in *Proc., IEEE Veh. Technology Conf. (VTC)*, Anchorage, Alaska, Sep. 2009.
- [18] H. Chao, Y. Chen, and J. Wu, "Power saving for machine to machine communications in cellular networks," in *Proc., IEEE Globecom Workshops*, Houston, TX, Dec. 2011, pp. 389 – 393.
- [19] A. Bartoli, J. Hernández-Serrano, M. Dohler, A. Kountouris, and D. Barthel, "Low-power low-rate goes long-range: The case for secure and cooperative machine-to-machine communications," in *Proc. Workshop on Wireless Cooperative Network Security (WCNS)*, Valencia, Spain, May 2011.
- [20] C.-Y. Tu, C.-Y. Ho, and C.-Y. Huang, "Energy-efficient algorithms and evaluations for massive access management in cellular based machine to machine communications," in *Proc., IEEE Veh. Technology Conf. (VTC)*, San Francisco, CA, Sep. 2011.
- [21] A. S. Lioumpas and A. Alexiou, "Uplink scheduling for machine-to-machine communications in LTE-based cellular systems," in *Proc., IEEE Globecom Workshops*, Houston, TX, Dec. 2011, pp. 353 – 357.
- [22] H. Viswanathan, "Uncoordinated orthogonal frequency division multiple access: To spread or not to spread," *IEEE Trans. on Wireless Communications*, vol. 8, no. 1, pp. 347 – 355, Jan. 2009.
- [23] B. Prabhakar, E. U. Biyikoglu, and A. El Gamal, "Energy-efficient transmission over a wireless link via lazy packet scheduling," in *Proc., IEEE INFOCOM*, Anchorage, Alaska, Apr. 2001, pp. 386 – 394.
- [24] A. El Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient scheduling of packet transmissions over wireless networks," in *Proc., IEEE INFOCOM*, New York, Jun. 2002.
- [25] I. Keslassy, M. Kodialam, and T. V. Lakshman, "Faster algorithms for minimum-energy scheduling of wireless data transmissions," in *Proc., Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, INRIA Sophia-Antipolis, France, Mar. 2003.
- [26] D. Slepian, "Bounds on communications," *The Bell System Technical Journal*, vol. 42, no. 3, pp. 681–707, May 1963.
- [27] M. J. Neely, "Intelligent packet dropping for optimal energy-delay tradeoffs in wireless downlinks," *IEEE Trans. on Automatic Control*, vol. 54, no. 3, pp. 565 – 579, Mar. 2009.
- [28] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.